National Climatic Data Center

DATA DOCUMENTATION

FOR

DATA SET 9641 (DSI-9641)

U.S. SNOW CLIMATOLOGY

December 17, 2002

National Climatic Data Center
151 Patton Ave.
Asheville, NC 28801-5001 USA

:

Table of Contents

:
:

1. **Abstract**: This ESDIM-funded project generated snow climatologies for 5525 stations in the contiguous United States and Alaska. These COOP station snow climatologies will be used to support NWS real-time snow operations in the ASOS observation era, and also will enable NOAA to better respond to user requests for snow information for use in economic and engineering decision-making.

This project analyzed daily snowfall and snow depth data from NCDC's DSI-3200 Cooperative Summary of the Day database. The digital period of record, through December 1996, was examined.

Several statistics were computed for several climatic elements using several snowfall and snow depth thresholds. The specific statistics computed vary with element. The statistics include mean, median, first and third quartiles, extremes, and probabilities. The elements include number of days with snow (snowfall or snow depth) beyond various thresholds, monthly and seasonal total snowfall, number of consecutive days with snow, dates of the first and last occurrence of snowfall, daily and multiple-day extreme snowfall amounts, and daily snow depth amount.

Several indicators based on the data and metadata were computed for each station to enable the user to assess the quality of the stations. Metadata are included for 24,646 stations in NCDC's Station History File, and snow data quality indicators are included for 11,525 stations in NCDC's DSI-3200 Summary of the Day digital data base.

Appendix 2-a has additional information on methodology.

2. **Element Names and Definitions**:

This data set consists of 8 files on 6 magnetic tapes. The first tape contains the station list file, station metadata file, and station QCI (quality control/inventory) file. The second tape contains the monthly/seasonal snow climatology statistics for stations in states Alabama through Missouri. The third tape contains the monthly/seasonal snow climatology statistics for stations in states Montana through Wyoming, plus Alaska. The fourth tape contains the daily snow climatology statistics for stations in states Alabama through Kansas. The fifth tape contains the daily snow climatology statistics for stations in states Kentucky through Ohio. The sixth tape contains the daily snow climatology statistics for stations in states Oklahoma through Wyoming, plus Alaska.

This ESDIM-funded project generated snow climatologies for 5525 stations in the contiguous United States and Alaska.

The files on magnetic tape 1 have the following specifications:

Record Length :	Fixed 165 characters
Blocked :	16500 characters
Media :	ASCII 18-Track IBM-Type 3480 cartridge
Parity :	Odd
Label :	ANSI Standard Labeled

The files on magnetic tapes 2 and 3 have the following specifications:

Record Length :	Fixed 121 characters
Blocked :	12100 characters

:
:

Media :	ASCII 18-Track IBM-Type 3480 cartridge
Parity :	Odd
Label :	ANSI Standard Labeled

The files on magnetic tapes 4, 5, and 6 have the following specifications:

Record Length :	Fixed 85 characters
Blocked :	8500 characters
Media :	ASCII 18-Track IBM-Type 3480 cartridge
Parity :	Odd
Label :	ANSI Standard Labeled

A. STATION LIST FILE (TAPE 1)

This file lists the stations (in ascending station number sort) used in the snow climatology project and has the following record format:

POSITION	WIDTH	TYPE	CONTENTS (Section 5 variable)
001-006	6	numeric	COOP station identification number (ICOOP)
007-007	1	alpha	not used
008-009	2	alpha	code indicating presence ('SF') or absence (' ') of snowfall statistics for this station (ASF)
010-010	1	alpha	not used
011-012	2	alpha	code indicating presence ('SD') or absence (' ') of snow depth statistics for this station (ASD)
013-165	153	alpha	not used

B. STATION METADATA FILE (TAPE 1)

The stations in NCDC's DSI-3200 Cooperative Summary of the Day data set were examined for quality and completeness. This examination included evaluation of the metadata in the Cooperative Network (COOP) station history, Historical Climatology Network (HCN) station history, and Climate Normals (CLIM81) historical time of observation (ob time) files, as well as evaluation of the data in the DSI-3200 data set. A number of indicators were generated for the stations and are included in two files:

(1) the station metadata file
(2) a quality control/inventory metadata file.

The station metadata file includes the following quality assessment indicators and standard location and identification information: number of years in the station history file, number of location changes, number of ob time changes, number of observer name changes, COOP station number, latitude, longitude, elevation, station and county names, state abbreviation, climate division, and metadata from the National Weather Service (NWS) Cooperative Substation Service Accountability (CSSA) data base. Blanks in some fields indicate that the metadata were not available. A code of -9 or -9.99 in some fields indicate the quality assessment indicator could not be computed or is not available/applicable for this station.

The station metadata are archived by ascending station number sort. Each station metadata record has the following record format:

POSITION WIDTH TYPE CONTENTS (Section 5 variable)

:

:

001-006	6	numeric	COOP station identification number (ICOOP)
001-006	6	numeric	COOP station identification number (ICOOP)
007-009	3	numeric	number of years station is in the COOP station history file (NYRC)
010-011	2	numeric	number of times station changed location, based on COOP station history file (NLOCC)
012-016	5	numeric	number of times station changed location, divided by number of years station is in the COOP station history file (CLOC)
017-018	2	numeric	number of times ob time changed, based on COOP station history file (NOBTC)
019-023	5	numeric	number of times ob time changed,divided by number of years station is in the COOP station history file (COBT)
024-026	3	numeric	number of years station is in the HCN station history file (NYRH)
027-028	2	numeric	number of times station changed location, based on HCN station history file (NLOCH)
029-033	5	numeric	number of times station changed location, divided by number of years station is in the HCN station history file (HLOC)
034-035	2	numeric	number of times ob time changed, based on HCN station history file (NOBTH)
036-040	5	numeric	number of times ob time changed, divided by number of years station is in the HCN station history file (HOBT)
041-042	2	numeric	number of times observer name changed, based on HCN station history file (NNAMEH)
043-047	5	numeric	number of times observer name changed, divided by number of years station is in the HCN station history file (HNAME)
048-049	2	numeric	number of years station is in the CLIM81 ob time file (NYRD)
050-051	2	numeric	number of times ob time changed, based on CLIM81 ob time file (NOBTD)
052-056	5	numeric	number of times ob time changed, divided by number of years station is in the CLIM81 ob time file (DOBT)
057-061	5	alpha-num	latitude (ddmmh, where dd=degrees, mm=minutes, h=hemisphere [N for North and S for South]) (LAT)
062-067	6	alpha-num	longitude (dddmmh, where ddd=degrees, mm=minutes, h=hemisphere [W for West, E for East]) (LON)
068-072	5	numeric	elevation, to whole feet (-999 = not available) (IELEV)
073-078	6	numeric	elevation, to meters and tenths (-999.9 = not available) (RELEV)
079-084	6	numeric	latest year-month (yyyymm) station is in COOP station history file (999999 = station open as of the end of 1996, 000000 = not known) (LYR,LMON)
085-085	1	alpha	not used
086-087	2	alpha	two-letter state abbreviation (STABBR)
088-088	1	alpha	not used
089-118	30	alpha-num	station name (SNAME)
119-120	2	alpha-num	climate division (values of 01-10, N, W, E, S) (DIV)
121-150	30	alpha	county name (CNAME)
151-155	5	numeric	Site I.D., from CSSA data base (SITEID)
156-159	4	numeric	the year that the current observer entered the

:

:

			Cooperative Program, from CSSA data base (CURYR)
160-160	1	alpha	Gender of the current observer, from CSSA data base (M=male, F=female, I=institution, U=government agency) (GENDER)
161-165	5	alpha	County Warning Area, from CSSA data base (WARN)

C. STATION QUALITY CONTROL/INVENTORY METADATA FILE (TAPE 1)

The stations in NCDC's TD-3200 Cooperative Summary of the Day data base were examined for quality and completeness. This examination included evaluation of the metadata in the Cooperative Network (COOP) station history, Historical Climatology Network (HCN) station history, and Climate Normals (CLIM81) historical time of observation (ob time) files, as well as evaluation of the data in the TD-3200 data
base. A number of indicators were generated for the stations and are included in two files: (1) the station metadata file, and (2) a quality control/inventory metadata file.

The quality control/inventory file includes the following quality assessment indicators: COOP station number, quality control (QC) indicators for snowfall and snow depth (number or percent of daily values suspect or converted to missing), and data inventory information (after QC) for snowfall and snow depth (digital period of record, number of years having data, number of months having complete data, number and percent of daily values missing, number of breaks). A code of -99.9 indicates a value that could not be computed.

The quality assessment indicators were generated based on data from 1948-1996. In the following table, the 'Contents' descriptions (for all fields except the first and last years of data in the data base, Positions 031-038) refer to this period.

The station metadata are archived by ascending station number sort, with one record for snowfall and one record for snow depth. Each station metadata record has the following record format:

POSITION	WIDTH	TYPE	CONTENTS (Section 5 variable)
001-006	6	numeric	COOP station identification number (ICOOP)
007-008	2	numeric	element code: 01 = snowfall 02 = snow depth (IEL)
009-013	5	numeric	number of non-missing daily values read from the DSI-3200 data base(NREAD)
014-018	5	numeric	number of daily values that were flagged as suspect by the quality control checks (NFLAG), including those flagged values that were corrected and those flagged values for which no corrective action was taken
019-023	5	numeric	number of daily values that failed the quality control checks and were set to missing (NMISS)
024-030	7	numeric	the percent of the non-missing daily values read that were flagged as suspect or set to missing 100%*[NFLAG+NMISS]/NREAD) (PCTMIS)
031-034	4	numeric	first year of data in DSI-3200 data base (IYRB)
035-038	4	numeric	last year of data in DSI-3200 data base (IYRE)
039-041	3	numeric	since data from only 1948-1996 were used to generate

:
:

			the QC and inventory indicators, this is the number of years between first and last years (if IYRB >= 1948), or number of years between 1948 and last year (if IYRB < 1948), or zero (if IYRE < 1948) (NYRDIF)
042-044	3	numeric	number of years in TD-3200 data base having some data (at least one day) (NYRSOM)
045-048	4	numeric	number of months having complete data (no days missing) (NMONC)
049-052	4	numeric	number of possible months in data period (number of months between first year first month of data and last year last month of data) (NMONP)
053-056	4	numeric	percent of possible months having complete data (100%*NMONC/NMONP) (PCTMON)
057-061	5	numeric	number of usable daily values processed (NPROC)
062-066	5	numeric	number of daily values missing (NMSG)
067-073	7	numeric	percent of daily values missing (100%*NMSG/[NMSG+NPROC]) (PCTDAY)
074-077	4	numeric	maximum number of consecutive whole months missing (biggest break) (MAXBRK)
078-080	3	numeric	number of breaks in record with any number of whole months missing (total number of breaks) (NBRKA)
081-083	3	numeric	number of breaks of different lengths (NBRKD)
084-165	82	alpha	not used

D. MONTHLY/SEASONAL SNOW CLIMATOLOGY STATISTICS DATA FILE (TAPES 2 AND 3)

The snow climatology monthly/seasonal output data are archived by ascending station number sort. Each data record consists of station identification information, data period, parameter information, twelve monthly data values, and six seasonal data values.

A five-tiered parameter coding system was developed to accommodate the large variety of parameters and statistics that were computed. The parameter coding system consists of a climatic element code, a statistic code, a date indicator code (which identifies the values as data values, dates, or years), a threshold code, and a time frame indicator code (which identifies the time frame as monthly/seasonal or daily). The codes are defined in Section 23 (Appendix A-1: Data Code Tables).

The data consist of decimal values, integer values, and special codes. Snowfall and snow depth amount are in inches. The special codes are defined as follows:

CODE VALUE	DEFINITION
-99.9	code for missing value due to data not available, insufficient data to compute a value, or statistic not applicable for this month or season
-99	code for missing value or statistic not applicable for this month or season
-66.6	code indicating value could not be computed due to insufficient data
-8.8	observed trace value

:
:

-7.7	code for computed value > 0.0 but < 0.5 (for snow depth) or 0.05 (for snowfall and dayswith), based on measurable observations

Each data record has the following record format:

POSITION	WIDTH	TYPE	CONTENTS (Section 5 variable)
001-006	6	numeric	Cooperative Network station number (ICOOP)
007-010	4	numeric	first year of input data analyzed (value of -99 indicates station never had measurable snowfall during the period analyzed, for the indicated element) (IYRB)
011-014	4	numeric	last year of input data analyzed (value of -99 indicates station never had measurable snowfall during the period analyzed, for the indicated element) (IYRE)
015-016	2	numeric	climatic element code (see Table 1) (IEL)
017-018	2	alpha	statistic code (see Table 2) (ASTAT)
019-019	1	alpha	date indicator code (see Table 3) (ADATE)
020-023	4	numeric	threshold code (see Table 4) (THRESH)
024-025	2	numeric	time frame indicator code (see Table 5) (ITIME)
026-030	5	numeric	January value (VALMON(1))
031-035	5	numeric	February value (VALMON(2))
036-040	5	numeric	March value (VALMON(3))
041-045	5	numeric	April value (VALMON(4))
046-050	5	numeric	May value (VALMON(5))
051-055	5	numeric	June value (VALMON(6))
056-060	5	numeric	July value (VALMON(7))
061-065	5	numeric	August value (VALMON(8))
066-070	5	numeric	September value (VALMON(9))
071-075	5	numeric	October value (VALMON(10))
076-080	5	numeric	November value (VALMON(11))
081-085	5	numeric	December value (VALMON(12))
086-091	6	numeric	value for Winter (Dec-Feb) (VALSEA(1))
092-097	6	numeric	value for Spring (Mar-May) (VALSEA(2))
098-103	6	numeric	value for Summer (Jun-Aug) (VALSEA(3))
104-109	6	numeric	value for Autumn (Sep-Nov) (VALSEA(4))
110-115	6	numeric	Annual (Jan-Dec) value (VALSEA(5))
116-121	6	numeric	value for snow season (Aug-Jul) (VALSEA(6))

E. DAILY SNOW CLIMATOLOGY STATISTICS DATA FILE (TAPES 4, 5, AND 6)

The snow climatology daily output data are archived by ascending station number sort. Each data record consists of station identification information, data period, parameter information, and twelve monthly data values.

A five-tiered parameter coding system was developed to accommodate the large variety of parameters and statistics that were computed. The parameter coding system consists of a climatic element code, a statistic code, a date indicator code (which identifies the values as data values, dates, or years), a threshold code, and a time frame indicator code (which identifies the time frame as monthly/seasonal or daily). The codes are defined in Section 23 (Appendix A-1: Data Code Tables).

The data consist of decimal values, integer values, and special codes. Snowfall and snow depth amount are in inches. The special codes are defined as

:

follows:

CODE VALUE	DEFINITION
-99.9	code for missing value due to data not available, insufficient data to compute a value, or statistic not applicable for this month or season
-99	code for missing value or statistic not applicable for this month or season
-66.6	code indicating value could not be computed due to insufficient data
-8.8	observed trace value
-7.7	code for computed value > 0.0 but < 0.5 (for snow depth) or 0.05 (for snowfall and days with), based on measurable observations

Each data record has the following record format:

POSITION	WIDTH	TYPE	CONTENTS (Section 5 variable)
001-006	6	numeric	Cooperative Network station number (ICOOP)
007-010	4	numeric	first year of input data analyzed (IYRB)
011-014	4	numeric	last year of input data analyzed (IYRE)
015-016	2	numeric	climatic element code (see Table 1) (IEL)
017-018	2	alpha	statistic code (see Table 2) (ASTAT)
019-019	1	alpha	date indicator code (see Table 3) (ADATE)
020-023	4	numeric	threshold code (see Table 4) (THRESH)
024-025	2	numeric	time frame indicator code (see Table 5) (ITIME)
026-030	5	numeric	January value (VALMON(1))
031-035	5	numeric	February value (VALMON(2))
036-040	5	numeric	March value (VALMON(3))
041-045	5	numeric	April value (VALMON(4))
046-050	5	numeric	May value (VALMON(5))
051-055	5	numeric	June value (VALMON(6))
056-060	5	numeric	July value (VALMON(7))
061-065	5	numeric	August value (VALMON(8))
066-070	5	numeric	September value (VALMON(9))
071-075	5	numeric	October value (VALMON(10))
076-080	5	numeric	November value (VALMON(11))
081-085	5	numeric	December value (VALMON(12))

3. Start Date: 18770101

4. Stop Date: 19961231

5. Coverage: North America

a. Southernmost Latitude: 24N
b. Northernmost Latitude: 72N
c. Westernmost Longitude: 174E
d. Easternmost Longitude: 67W

:

6. How to Order Data:

Ask NCDC's Climate Services about the cost of obtaining this data set.
Phone: 828-271-4800
FAX: 828-271-4876
E-mail: **NCDC.Orders@noaa.gov**

7. Archiving Data Center:

National Climatic Data Center
Federal Building
151 Patton Avenue
Asheville, NC 28801-5001
Phone: (828) 271-4800.

8. Technical Contact:

National Climatic Data Center
Federal Building
151 Patton Avenue
Asheville, NC 28801-5001
Phone: (828) 271-4800.

9. Known Uncorrected Problems: The properties of snow make it difficult to accurately and consistently measure snowfall. Snow often melts as it lands or as it lies on the ground and settles as it lies on the ground. Additionally, snow is easily blown and redistributed. These properties can be affected by location, the time of day the snow observations are taken, and how often they are measured (Doesken and Judson, 1997). For these reasons, it is important for observers to adhere to a standard methodology (National Weather Service, 1972) for observing and reporting snow. Unfortunately, stations change location, observers, and sometimes observation time. Such changes introduce inhomogeneities into the snow record. No acceptable adjustment algorithms exist to statistically adjust daily snow data for these inhomogeneities.

10. Quality Statement: A quality control methodology was employed to correct or eliminate snowfall and snow depth values that were determined to be in error. However, as noted in paragraph 9 above, (Known Uncorrected Problems), the data were not adjusted for inhomogeneities. Several indicators were generated from the data and metadata to allow the user to estimate the degree of confidence that could be placed in the statistics for each station. The quality control methodology and QC/inventory indicators are described in Appendix A-2: Project Methodology.

11. Essential Companion Datasets: None.

12. References:

Doesken, N.J. and A. Judson, 1997: A Guide to the Science, Climatology, and Measurement of Snow in the United States, Second Edition, Colorado State University Department of Atmospheric Science: Fort Collins.

Easterling, D.R., T.R. Karl, E.H. Mason, P.Y. Hughes, and D.P. Bowman (R.C. Daniels and T.A. Boden, editors), 1993: United States Historical Climatology Network (U.S. HCN): Monthly Temperature and Precipitation Data. Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Publication No. 4500, ORNL/CDIAC-87, NDP-019/R3, Oak Ridge, TN.

:
:

Ludlum, D.M., 1982: The American Weather Book, Houghton Mifflin Co.: Boston.

National Weather Service, 1972: National Weather Service Observing Handbook No. 2: Substation Observations, First Edition, Revised December 1972 (Supersedes Circular B), U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Reek, T., S.R. Doty, and T.W. Owen, 1992: AA deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network.@ Bulletin of the American Meteorological Society, vol. 73, pp. 753-762.

:

Appendix A-1: Data Code Tables

Table 1. Climatic element codes and description.

Code	Description
01	snowfall
02	snow depth (snow cover)
10	number of days with daily snowfall amount equal to the specified threshold
11	number of days with daily snowfall amount greater than or equal to the specified threshold
12	number of two-day periods with two-day snowfall amount equal to the specified threshold
13	number of two-day periods with two-day snowfall amount greater than or equal to the specified threshold
14	number of three-day periods with three-day snowfall amount equal to the specified threshold
15	number of three-day periods with three-day snowfall amount greater than or equal to the specified threshold
20	monthly or seasonal total snowfall amount
30	number of consecutive days with daily snowfall amount greater than or equal to the specified threshold
38	first occurrence of daily snowfall amount greater than or equal to the specified threshold
39	last occurrence of daily snowfall amount greater than or equal to the specified threshold
40	daily snowfall amount (all days examined, whether they had snowfall or not)
41	daily snowfall amount (only days having snowfall examined)
42	greatest two-day total snowfall amount (where snow fell on each day) in a month
43	greatest three-day total snowfall amount (where snow fell on each day) in a month
44	greatest four-day total snowfall amount (where snow fell on each day) in a month
45	greatest five-day total snowfall amount (where snow fell on each day) in a month
46	greatest six-day total snowfall amount (where snow fell on each day) in a month
47	greatest seven-day total snowfall amount (where snow fell on each day) in a month
48	greatest one-day snowfall amount in a month
50	number of days with snow depth equal to the specified threshold
51	number of days with snow depth greater than or equal to the specified threshold
70	number of consecutive days with snow depth greater than or equal to the specified threshold
71	greatest number of consecutive days with snow depth greater than or equal to the specified threshold
80	daily snow depth amount (all days examined, whether they had snow cover or not)
81	daily snow depth amount (only days with snow cover examined)
92	greatest two-day total snowfall amount (whether it snowed each day or not) in a month
93	greatest three-day total snowfall amount (whether it snowed each day or not) in a month
:	
:	

Table 2. Statistic codes and description.

Code	Description
DM	median date in the month
ER	earliest date in the month
G1	greatest value (same as MX)
G2	2nd greatest value
G3	3rd greatest value
G4	4th greatest value
G5	5th greatest value
G6	6th greatest value
G7	7th greatest value
G8	8th greatest value
G9	9th greatest value
G0	10th greatest value
LA	latest date in the month
MD	median (50th percentile)
MI	minimum (same as S1)
MN	average (mean)
MX	maximum (same as G1)
ND	number of years with daily snow depth greater than or equal to the specified threshold
NF	number of years with daily snowfall greater than or equal to the specified threshold
NY	number of years with non-missing data
PR	probability (of receiving measurable snowfall) times 100%
Q1	first quartile (25th percentile)
Q3	third quartile (75th percentile)
S1	smallest value (same as MI)
S2	2nd smallest value
S3	3rd smallest value
S4	4th smallest value
S5	5th smallest value
S6	6th smallest value
S7	7th smallest value
S8	8th smallest value
S9	9th smallest value
S0	10th smallest value

Table 3. Date indicator codes and description.

Code	Description
blank	blank indicates the value is a data value
D	value is a date: for January-December, the value is the day of the month; for the seasons (winter, spring, summer, autumn, annual, and Aug-Jul season), the value is the month (MM) and day (DD) coded as MMDD
Y	value is a year

:

Table 4. Threshold codes and description.

Code	Description
####	a snowfall or snow depth threshold value in inches and tenths
-8.8	a trace value (a value > 0.0 inches but < 0.05 inches for snowfall, or a value > 0.0 inches but < 0.5 inches for snow depth)
-9.9	threshold does not apply to this record

Table 5. Time frame indicator codes and description.

Code	Description
##	a value for ## of 01-31 indicates the day of the month the daily value is for
00	the value is for a month or season

Appendix A-2: Project Methodology

1. Production of Snowfall and Snow Depth Climatologies for NWS Cooperative Observer Sites

1.1 Objective

The purpose of this ESDIM-funded project was to generate snowfall and snow depth statistics for several thousand non-airport stations in the National Weather Service (NWS) Cooperative (COOP) Network. Stations in the 49 continental states were considered. The Automated Surface Observing System (ASOS) instrumentation being installed at airport locations detects weather phenomena, including the occurrence of snow, using standard observing methodology. However, the ASOS automated instruments are not able to measure the amount of snowfall or snow on the ground (snow depth). This project established snow climatologies for COOP stations that could be used to support NWS real-time snow operations in the ASOS observation era. These snow climatologies also enable the National Oceanic and Atmospheric Administration (NOAA) to better respond to user requests for snow information for use in economic and engineering decision-making.

1.2 Data

This project analyzed daily snowfall and snow depth data from NCDC’s TD-3200 Cooperative Summary of the Day database. The digital period of record, through December 1996, was examined. Daily maximum and minimum temperature and precipitation were used to quality control (QC) the snow data.

1.3 Quality Control

Three levels of quality control were employed in order to obtain the best snow data possible. The first level involved using the ValHiDD edited TD-3200 values. The second level employed a number of internal consistency checks. The third level was an extremes check.

1.3.1 First Level QC: ValHiDD

During the 1990's, an automated quality control system called ValHiDD (Validation of Historical Daily Data) was applied to the entire DSI-3200 data set to remove gross errors in daily maximum and minimum temperature, precipitation, snowfall, and snow depth. ValHiDD is a rules-based method for detecting and correcting discrepancies (due to digitizing errors and observer errors) in the DSI-3200 data set. The checks employed by ValHiDD include a limits check, internal consistency checks, flatliner temperature check, precipitation/snowfall/snow depth (PSFSD) relationship check, temperature spike check, multiple rule-group failures check, and failed fix check (Reek, et al., 1992).

Although the number of discrepancies uncovered and resolved by ValHiDD was small compared to the total number of data values examined, their removal/correction was important and ValHiDD significantly contributed to the improvement of the overall DSI-3200 data set. However, two factors relevant to this project should be noted:

(1) In some PSFSD cases, ValHiDD could not identify which element should be corrected, so the values were flagged as suspect and not altered.

:

(2) The PSFSD relationship check assumed that all three elements were observed at the same hour. For most volunteer COOP observers this assumption holds. However, for airport stations, this is not the case: snow depth is observed at 7 a.m. local time, while daily snowfall and precipitation amount are reported as of midnight. This airport station observation time discrepancy complicated the PSFSD relationship check.

1.3.2 Second Level QC: Internal Consistency Checks

This level of QC included temporal checks (comparing today's snow depth values to yesterday's values) and additional inter-element checks beyond those performed by ValHiDD. Snowfall and snow depth values that failed the internal consistency checks were corrected (where possible) or set to missing. Temperature or precipitation values were not examined for accuracy at this level.

The second level QC included the following checks. The following abbreviations are used here: TMIN = minimum temperature (deg. F), TMAX = maximum temperature (deg. F), P = precipitation (inches), SF = snowfall (inches), and SD = snow depth (inches).

(1) Factor of 10 error for SF: if P >= 0.01 and SF >= 1.0 and the ratio, SF/P, was greater than 80.0, then the SF value was corrected by dividing by 10. The corrected SF value was similarly checked and set to missing if the new SF/P ratio was greater than 50.

(2) Hail check: non-zero SF values were set to zero if TMIN >= 40.

An alternative check was used for those cases where the minimum temperature was missing (stations measuring both temperature and precipitation where the day's TMIN was missing, and stations which measured only precipitation). This alternative involved examining the day's climatological median extreme minimum temperature (CMEMT) for the state. The CMEMT was computed for each of the 365 days of the year (the value for February 28 was used for February 29 leap days) for each state from the daily extreme minimum temperature values for all stations in the state, from the period 1961-1990. Non-zero SF values were set to zero if CMEMT > 25.

(3) Non-zero SF values were set to missing if:
- SF > 0.4 but P = 0; or
- today's P is missing.

(4) Factor of 10 error for SD (where previous day's SD = zero or trace): SD was compared to SF and corrected if it was identified as being off by a factor of 10. If the SD was greater than ten times SF, the SD was set to missing. (There were a few cases where the observer inconsistently recorded SD off by a factor of ten for a string of years. This check was used to identify the beginning and ending years of such periods, so the station's data could be later examined manually. If the error was not consistent, the snow depth from this string of years was subsequently deleted from the analysis.)

(5) Second check for factor of 10 error for SD: if the difference between today's SD and yesterday's SD was greater than today's SF (plus an adjustment factor due to difference in units resolution), today's SD was divided by 10. The corrected SD value was similarly checked and set to missing if the difference in SD was still greater than today's SF.

(6) Zero SD values were set to missing if:
yesterday’s SD > 7 and today’s SF > 2.0.

(7) Non-zero SD values were set to missing if:
- today’s SD > yesterday’s SD with today’s SF = 0; or
- today’s SF is missing; or
- yesterday’s SD missing and today’s SD > (today’s SF + SD of last day with non-missing SD).

(8) Today’s SD was set to missing if today’s P < 0.05 and:
- yesterday’s SD >= (4 + today’s SD), and today’s TMAX < 40; or
- yesterday’s SD >= (7 + today’s SD), and today’s TMAX < 45; or
- yesterday’s SD >= (10 + today’s SD), and today’s TMAX > 44; or
- yesterday’s SD >= (7 + today’s SD), and today’s TMAX missing.

The snowfall and snow depth values that were corrected or set to missing (by the above 8 checks) were tallied and the counts were saved to a metadata file to be used later in a station quality assessment step.

The following additional checks were made. Values failing these checks were not changed, but the number of flagged values was similarly saved to a metadata file.

(9) Questionable SF values (the SF/P ratios were unusual):
- 1 < SF < 3, and SF > 50*P; or
- 3 <= SF <= 6, and SF > 40*P, and TMAX > 24; or
- SF > 6, and SF > 20*P, and TMAX > 24; or
- SF > 6, and SF > 30*P, and TMAX < 25.

(10) Questionable SD values (unusual decrease in SD):
- today’s SD <= yesterday’s SD, today’s SF > 2.0, and today’s TMAX < 30.

As in the ValHiDD discussion (see section 1.3.1), it is critical to these automated tests that the temperature, precipitation, snowfall, and snow depth observations be taken at the same hour. This is the case for most volunteer COOP observers. However, NWS and Federal Aviation Administration (FAA) airport stations observe snow depth at 7 a.m. local time, while the remaining elements are reported as of midnight. This airport station observation time discrepancy impacts checks (4)-(7) and (10) above, and can result in valid snow depth values being flagged as erroneous and being changed.

1.3.3 Third Level QC: Extremes Checks

The daily snowfall values were compared, on a state-by-state basis, to known statewide 24-hour snowfall extremes. The known extremes published by weather historian David Ludlum (Ludlum, 1982) were, for most states, multiplied by an acceptability factor (1.4) in order to account for new daily extremes that may have been set since his book was published, and to account for the difference in time frame (a moving 24-hour time frame versus daily values taken at a fixed ob time). Special subjective estimates were used for Colorado, Florida, and New York.

These adjusted statewide extremes were used in the snowfall extremes check. If a station’s daily snowfall value exceeded the corresponding statewide extreme, the value was set to missing and the occurrence was tallied. The counts were saved to a metadata file to be used later in a station quality assessment step.

:
:

Snow depth varies widely in states with mountain topography. For example, the extremes for coastal stations in southern California would be considerably lower than the extremes for stations in the Sierra Nevada range. This made it difficult to establish an appropriate statewide snow depth extreme, so a standard snow depth extreme of 2000 inches was used for all stations in all 49 states. If a station's snow depth value exceeded 2000 inches, the value was set to missing and the occurrence was similarly tallied and saved.

1.4 Analysis Procedures

The NWS Office of Meteorology (OM) solicited input on suggested methodology and desired output statistics from NWS regional and field offices and from non-NOAA snow experts. This input was reviewed by NCDC and NWS OM and incorporated, as appropriate, into the project.

The properties of snow make it difficult to accurately and consistently measure snowfall and snow depth. Snow often melts as it lands or as it lies on the ground, snow settles as it lies on the ground, and snow is easily blown and redistributed. These properties can be affected by location, time of day the observations are taken, and how often they are measured (Doesken and Judson, 1997). For these reasons, it is important for observers to adhere to a standard methodology (National Weather Service, 1972) for observing and reporting snow. Unfortunately, stations change location, observers, and sometimes observation time. Such changes introduce inhomogeneities into the snow record. No acceptable adjustment algorithms exist to statistically adjust daily snow data for inhomogeneities. The alternative for creating a reasonably high quality set of snow statistics, therefore, is to use stations which have a low risk of having inhomogeneous data.

For this project, the entire DSI-3200 data set was examined. QC (Section 1.4.1) and inventory (Section 1.4.2) indicators were computed for data from the period 1948-1996. Several station metadata files were examined and metadata indicators were computed (Section 1.4.3). The QC, inventory, and station metadata indicators were used to assess the quality of each station (Section 1.4.4). The stations included in the final station list were selected based upon this objective assessment of their quality, as well as (where human resources allowed) a subjective assessment (based on experience with operational processing of the stations= data).

Two sets of QC and inventory indicators were computed for each station, one for snowfall and one for snow depth. As a result, some stations will have output products and statistics for snowfall but not snow depth, some for snow depth but not for snowfall, and some for both snowfall and snow depth.

1.4.1 Data QC Indicators

Daily COOP data have been digitized operationally beginning in 1948. Over the years, pre-1948 data have been keyed on a special project case-by-case basis, and are more likely to have gaps of missing data. About 18 percent of the stations had data beginning before 1948, while less than one percent (0.36%) had data that ended before 1948. Consequently, the QC indicators were computed for the 1948-1996 period (however, the entire data base was QC=d and analyzed for the computation of the statistics). The QC indicators include the following:

(1) the number of non-missing daily values read;

:

:

(2) the number of daily values that were flagged as suspect by the QC checks, including those flagged values that were corrected and those flagged values for which no corrective action was taken;

(3) the number of daily values that failed the quality control checks and were set to missing; and

(4) the percent of the non-missing daily values read that were flagged as suspect or set to missing.

1.4.2 Data Inventory Indicators

For the reasons noted in Section 1.4.1, the inventory indicators were computed over the period 1948-1996. They include the following:

(1) number of years in the TD-3200 data base between the first and last years with data;

(2) number of years in the TD-3200 data base having some data (at least one day);

(3) number of months having complete data (no days missing), and percent of possible months having complete data;

(4) number of usable daily values processed;

(5) number of daily values missing, and percent of daily values missing; and

(6) information concerning the number of breaks (or gaps) in the data record, where a break is defined as at least one month completely missing. The information included the number of breaks of different lengths, the total number of breaks (breaks with any number of months missing), and the length of the biggest break. For example, if a station had one break of three months length and two breaks of five months length, then it would have two breaks of different lengths, three breaks in total, and the biggest break would be five months long.

1.4.3 Station Metadata Indicators

Station metadata from three files were utilized in the creation of the metadata indicators: the Cooperative Station History File (COOP); the U.S. Historical Climatology Network Station History File (HCN); and the 1961-1990 Climate Normals historical time of observation (ob time) file (CLIM81). The COOP file contains metadata for all (approximately 24,000) stations, past and present, in the Cooperative Network, but some metadata elements are not complete. The HCN file contains complete metadata for the 1221 Abest@ stations in the Cooperative Network (Easterling, et al., 1996). The CLIM81 file contains complete ob time metadata for the 40 years, 1951-1990, for the 6662 stations for which 1961-1990 monthly climate normals were computed. Metadata from these three files were examined in order to capture the most comprehensive metadata information for as many stations as possible.

The following station metadata indicators were computed:

(1) from the COOP file: number of years the station is in the file, the number of times the station changed location, and the number of times the station

:

changed location divided by how long it is in the file.

Location is measured in the COOP file by latitude, longitude, elevation, and a Arelocation@ (station moved x distance in y direction) parameter. Latitude, longitude, and elevation information was available for the period of record, however relocation information was not available for 1948-1980.

Some tolerance was built into this indicator. A location change occurred if any of the following criteria were met:

a. any change in latitude or longitude (both measured to the nearest minute);

b. an elevation change greater than 20 feet; or

c. the relocation parameter indicated a move greater than one tenth of a mile. Of the 5631 decipherable relocations in the metadata base, approximately 26% of them were less than or equal to a tenth of a mile.

(2) from the COOP file: the number of times the station's ob time changed, and the number of times the station's ob time changed divided by how long it is in the file.

Ob time information is available from only 1981-present. There are three pronounced peaks in a plot of ob time change: one at 1 hour, one at 9-10 hours, and one at 16-17 hours. One can safely assume that an ob time change of 1 or 2 hours will not introduce a significant inhomogeneity into the snow record. Ob time changes of 3-5 hours are rare. For these reasons, some tolerance was built into this indicator, as well, with an ob time change being counted only if it exceeded 2 hours.

(3) from the HCN file: number of years the station is in the file, the number of times the station changed location, and the number of times the station changed location divided by how long it is in the file.

The discussion for the COOP location indicator applies to the HCN location indicator, except the HCN location information was available for the entire period of record.

(4) from the HCN file: the number of times the station's ob time changed, and the number of times the station's ob time changed divided by how long it is in the file.

The discussion for the COOP ob time change indicator applies to the HCN ob time change indicator, except the HCN ob time information was available for the entire period of record.

(5) from the HCN file: the number of times the station's observer name changed, and the number of times the station's observer name changed divided by how long it is in the file.

(6) from the CLIM81 file: number of years the station is in the file, the number of times the station's ob time changed, and the number of times the station's ob time changed divided by how long it is in the file.

The discussion for the COOP ob time change indicator applies to the CLIM81 ob time change indicator, except the CLIM81 ob time information was available for

:

the CLIM81 file's period of record.

1.4.4 Station Quality Assessment

The following QC, inventory, and station metadata indicators were used to assess the quality of each station. Examination of frequency distribution charts of these indicators did not provide meaningful guidance in determination of cutoff criteria. Therefore, the specific criteria chosen were selected in order to maximize both the quality of the station data set and the number of stations included in the data set. The data and metadata indicators for all stations are included in the metadata files, in the event the user wishes to apply different criteria.

In order to be included in the project's final station list (for snowfall and/or snow depth), the station had to meet the following requirements:

(1) have at least 15 years of non-missing data for each of the 12 months (January-December) for selected climatic elements (number of days with snowfall, monthly total snowfall amount, greatest daily snowfall amount, number of days with snow depth, and daily snow depth amount);

(2) have at least 15 years of non-missing data for each of the 365 days of the year for selected climatic elements (daily snowfall amount and daily snow depth amount);

(3) have at least 70% of the months from the data period with complete data (no days missing);

(4) have 33 or fewer breaks per 100 years;

(5) have no more than 25% of the daily values missing out of the total number of values with data>;

(6) have 3 or fewer ob time changes, based on the COOP metadata file;

(7) have 10 or fewer location changes per 100 years, based on the COOP metadata file (it should be noted that latitude, longitude, and/or elevation may have changed due to the switch from manually-based surveys to satellite-based surveys when, in fact, the station did not move); and

(8) have 10 or fewer ob time changes per 100 years, based on the CLIM81 metadata file.

Stations which met these criteria were then examined for station type. NWS offices and NWS and FAA airport stations were deleted from the snow depth list, due to QC considerations as discussed in sections 1.3.1 and 1.3.2.

1.5 Computational Methodology

A suite of statistics (mean, median, first and third quartiles, extremes [both amounts and dates of occurrence], and frequencies/probabilities) was generated for several climatic parameters. The specific statistics that were computed vary with parameter, but the number of years with non-missing data (NYRS) was computed for each parameter. The beginning and ending years and NYRS information are crucial for any inter-station or inter-seasonal comparisons the user may wish to make.

:

A daily climatology and a monthly/seasonal climatology were created for each station. The statistics for the daily climatology were generated for each day from the years of data available for the day. The statistics for the monthly/seasonal climatology were generated from year-month or year-season sequential values.

1.5.1 Climatic Parameters

The climatic parameters include the following:

(1) number of days (1-day periods, 2-day periods, and 3-day periods) with daily snowfall amount equal to zero or a trace;

(2) number of days (1-day periods, 2-day periods, and 3-day periods) with daily snowfall amount greater than or equal to several thresholds (0.1, 1.0, 2.0, 5.0, 10.0, 12.0, 18.0, 24.0, and 36.0 inches);

(3) monthly and seasonal total snowfall amount;

(4) number of consecutive days with daily snowfall amount greater than or equal to several thresholds (0.1, 1.0, 2.0, and 5.0 inches);

(5) dates of first and last occurrence if daily snowfall amount greater than or equal to several thresholds (1.0, 4.0, and 6.0 inches);

(6) daily snowfall amount, both with all days examined (whether they had snowfall or not) and only days having snowfall examined;

(7) greatest multiple-day (2-, 3-, 4-, 5-, 6-, and 7-day) total snowfall amount (where snow fell on each day) in a month;

(8) greatest 2-day and 3-day total snowfall amount (whether it snowed each day or not) in a month;

(9) number of days with snow depth equal to zero or a trace;

(10) number of days with snow depth greater than or equal to several thresholds (1.0, 2.0, 5.0, and 10.0 inches);

(11) number of consecutive days with snow depth greater than or equal to various thresholds (1.0, 2.0, 5.0, 10.0, 12.0, 18.0, 24.0, and 36.0 inches);

(12) daily snow depth amount, both with all days examined (whether they had snow cover or not) and with only days with snow cover examined; and

(13) number of years (from which frequencies were derived) with snowfall or snow depth greater than or equal to several thresholds (0.1, 1.0, 2.0, 5.0, 10.0, and 12.0 inches) for each day of the year.

1.5.2 Monthly/Seasonal Climatology Computational Considerations

The multiple-day extremes parameters (see group 7 in Section 1.5.1) are based on snow falling on each of the days in the time unit. If snow fell on some days but not all days, then that value was not included in the analysis for that multiple-day time unit, but it may qualify for inclusion in a shorter time unit. This could result in some longer time units having smaller extreme snowfall amounts than the corresponding shorter time units.

:

The date (year, month, and/or day) of an extreme is the date of the most recent occurrence (except for statistics G1-G0). The date listed for multiple-day parameters is the last day of the multiple-day period.

Monthly statistics (for January through December) were computed based on the days in the month under consideration. Seasonal statistics were computed for winter, spring, summer, autumn, annual, and snow season, with the seasons corresponding to the following months, respectively:

December-February, March-May, June-August, September-November, January-December, and August-July. The seasonal statistics are not based on the monthly statistics; they were computed from the daily values corresponding to each season in each year of the record. (For example, the mean winter statistics are not the average or total of the December, January, and February mean statistics; they are based on the sequential winter periods in the data record.) For the first and last occurrence of snowfall (group 5 in Section 1.5.1), the (incomplete) years at the beginning and end of the data period were included in the analysis for the seasonal statistics. For these reasons, the seasonal statistics may not agree with the corresponding monthly statistics.

For the first and last occurrence of snowfall (group 5 in Section 1.5.1), if no snow occurred, then there was no data from which to compute the dates and the first and last years of data will be coded as Amissing@ (-99). For these elements, the NYRS statistics refers to the number of years with non-zero data, which is somewhat broader than ‘the number of years with non-missing data.’

Likewise for the daily snowfall (snow depth) amount, where only days having snowfall (snow depth) were examined (groups 6 and 12 in Section 1.5.1): if no snow occurred, then there was no data from which to compute a value. Consequently, the NYRS statistics refers to the number of years with non-zero data, which is somewhat broader than ‘the number of years with non-missing data.’

The greatest number of consecutive days with snow depth parameter (element 71) was computed for just the August-July snow season. It was not computed for the other seasons or the individual months.

1.5.3 The Effect of Missing Data

The impact of missing data varies, depending on the element and statistic computed. Total snowfall amount had no tolerance for missing data. If even one day was missing in a month or season, the total snowfall could not be computed for that year’s month or season. Consequently, the number of years with non-missing data will vary with month and season. The six seasons (especially annual and August-July) have a greater chance of experiencing missing data and, generally, will have fewer years with non-missing data when compared to the individual months.

The median daily value for a month had no tolerance for missing data. All days in a month had to have data in order for a median daily value to be computed for that year-month.

The number of days with snowfall or snow depth parameters had no tolerance for missing data. Data for leap days (February 29) were included in the analysis.

:

Due to this fact and due to rounding error, the sum of the values for the equal zero, equal trace, and greater than or equal to 0.1 inch (for snowfall, 1.0 inch for snow depth) thresholds may not exactly equal the maximum possible number of days in the month or season. This will be especially noticeable for the number of days with 2-day and 3-day snowfall parameters.

The consecutive days with snowfall or snow depth, or 'runs' parameters, had no tolerance for missing data for each specific threshold. A run of consecutive days for a given threshold was delineated by days (immediately before and after the run) that had values less than the run's threshold value. Consequently, a run had to have no missing days during the run and on the days immediately before the run started and after the run ended in order to be included in the analysis. If a non-zero day outside of this range was missing, however, then the runs for the lower thresholds would be affected and would be treated as missing in the sequential data for that year. This could result in statistics for runs with higher thresholds being larger than the corresponding statistics for lower thresholds.

The daily extreme, multiple-day extreme, and date of occurrence parameters had a greater tolerance for missing data. Data were analyzed even if a month had up to 5 days missing. This could result in apparent discrepancies between these and other parameters.

The greatest 2-day and 3-day total snowfall amount, whether it snowed each day or not (group 8 in Section 1.5.1), could tolerate up to 5 missing days per month. However, missing days within a 2- or 3-day period were excluded from the analysis. (For example, if a 2-day period had snow, but the day before and the day after this 2-day period were missing, then the snowfall total would be included in the 2-day analysis but not in the 3-day analysis.) Missing days might, in some cases, result in extreme 2-day snowfall totals being greater than the corresponding extreme 3-day snowfall totals.

A complicated tolerance for missing data was built into the probability of receiving measurable snowfall parameter (statistic PR). The data in a month (or season) were examined to determine if it snowed in a given year-month (or year-season). If even one day in a month (or season) had measurable snowfall, then that year was counted in the computations, regardless of how many days were missing. However, to determine if no snow occurred in a month (or season), the month (season) had to have no days missing. The probability for each month (season) was computed by summing the number of years with one or more days of measurable snowfall, then dividing by the number of years that qualified (i.e., where it could be determined that it did or did not snow).

1.6 References

Doesken, N.J. and A. Judson, 1997: A Guide to the Science, Climatology, and Measurement of Snow in the United States, Second Edition, Colorado State University Department of Atmospheric Science: Fort Collins.

Easterling, D.R., T.R. Karl, E.H. Mason, P.Y. Hughes, and D.P. Bowman (R.C. Daniels and T.A. Boden, editors), 1993: United States Historical Climatology Network (U.S. HCN): Monthly Temperature and Precipitation Data. Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Publication No. 4500, ORNL/CDIAC-87, NDP-019/R3, Oak Ridge, TN.

Ludlum, D.M., 1982: The American Weather Book, Houghton Mifflin Co.: Boston.

National Weather Service, 1972: National Weather Service Observing Handbook No. 2: Substation Observations, First Edition, Revised December 1972 (Supersedes Circular B), U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Reek, T., S.R. Doty, and T.W. Owen, 1992: 'A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network.' Bulletin of the American Meteorological Society, vol. 73, pp. 753-762.

1.7 Acknowledgments

This project was supported by ESDIM grant NWS-97-04. The project leaders would like to specially thank Dr. David Robinson of Rutgers University, Mr. Nolan Doesken of the Colorado Climate Center, and Mr. Grant Goodge of Asheville, NC for their valuable input and suggestions.

:
: